DOCUMENT RESUME

ED 308 211 TM 013 528

AUTHOR	Linacre, John M.
TITLE	Rasch Models from Objectivity: A Generalization.
PUB DATE	Mar 89
NOTE	11p.; Paper presented at the International Objective Measurement Workshop (5th, Berkeley, CA, March 26, 1989).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Equations (Mathematics); Evaluators; *Generalization; *Latent Trait Theory; Mathematical Models; *Psychometrics; *Rating Scales
IDENTIFIERS	*Objectivity; Paper and Pencil Tests; Par 'etric Analysis; *Rasch Model; Three Parameter Model; Two Parameter Model

ABSTRACT

Rasch models for fundamental measurement in the psychological sciences are derived from the principle of specific objectivity, the requirement that the parameter value representing each component in a test situation be independent of the other components. The dichotomous Rasch model for two-faceted analysis, applicable to conventional paper-and-pencil tests, is constructed. The many-faceted Rasch model is also derived by means of a three-faceted example, comprising judges, examinees, test items, and a rating scale, which is applicable to many judging situations. Any other particular form of the many-faceted model may also be derived in a similar manner. (Author/TJH)

ED308211

Rasch Models from Objectivity:

A Generalization

John M. Linacre

M E S A

University of Chicago

5835 S. Kimbark Avenue

Chicago Il 60637

Paper presented at the
Fifth International Objective Measurement Workshop
University of California, Berkeley, CA March 26, 1989.

Abstract:

Rasch models for fundamental measurement in the psychological sciences are derived from the principle of specific objectivity, the requirement that the parameter value representing each component in a test situation be independent of the other components. The dichotomous Rasch model for two-faceted analysis, applicable to conventional paper-and-pencil tests, is constructed. The many-faceted Rasch model is also derived by means of a three-faceted example, comprising judges, examinees, test items and a rating scale, which is applicable to many judging situations.

Key-words: Rasch measurement, Objectivity, Rating scales

I. Introduction.

Georg Rasch has some wise words to say on the subject of objectivity which will guide us in this discussion:

"The concept of 'objectivity' raises fundamental problems in all sciences. For a statement to be scientific, 'objectivity' is required. However, exactly what 'objectivity' means is disputed among philosophers and I am not going to enter into that debate" (Rasch 1964 p.1).

"The comparison of any two subjects may be carried out in such a way that no other parameters are involved than those of the two subjects - neither the parameter of any other subject nor of any of the stimulus parameters.

"Similarly, any two stimuli may be compared independently of all other parameters than just those of the two stimuli - the parameters of all other stimuli as well as the parameters of the subjects having been replaced by observable numbers.

"It is suggested that comparisons carried out under such circumstances will be designated as 'specifically objective'. And the same term would seem appropriate for statements about the model structure which are independent of all of the parameters specified in the model, the unknown values of them being, in fact, irrelevant for the structure of the model" (Rasch 1966 p.21).

"A model is not meant to be true. Even in classical physics models are temporary - good enough for some purposes" (Rasch 1964 p. 2).

Our aim, then, is to characterize each component of a complex test situation by one parameter of fixed, but unknown, value which is objective, that is to say independent of the other parameters. These parameters are to be estimated, or measured, by means of a model which capitalizes on this invariance in the parameters.

II. Objectivity for a two-faceted test.

Rasch developed the idea of objectivity in his analysis of an intelligence test taken by 1094 recruits to the Danish Army. (Rasch 1980 p.62). This was a conventional two-faceted test situation: the observations were the dichotomous (right-wrong, 1-0) results of interactions between the objects (examinees) and the agents (test items).

The derivation of the corresponding Rasch model, from objectivity, is included here as a demonstration of the principles involved following Wright (passim, in particular Wright and Linacre, 1987).

Consider a test in which two objects, Om and On, respond to numerous replications of a dichotomously scored agent, Ai. The outcome of this test is depicted in Figure 1.

	Object On 0	Object On 1
Object Om 0	F00	F10
1	F01	F11

Figure 1. Outcome of the hypothetical administration of numerous replications of an agent to two objects. F10 is the count of the number of times that object On succeeded on the agent at the same time that object Om failed.

The only way to determine the difference between two objects is to observe when they perform differently. In Figure 1, they have performed in the same way with counts of F00 and F11; they have performed differently with counts of F10 and F01. The comparison of On and Om is thus observed through the comparison of F10 and F01, for which F00 and F11 provide a quantitative context. In whatever way the comparison of F10 and F01 is made, it must be independent of the length of the test. Thus, if the test were to be twice as long, with identically proportionate results, then the comparison between Om and On would become a comparison between 2*F10 and 2*F01. In order for this comparison to be independent of the arbitrary number of replications, the comparison between Om and On can be estimated by

$$\frac{\text{Performance of On}}{\text{Performance of Om}} \approx \frac{F10}{F01} \qquad (1)$$

This ratio of frequencies is the empirical but stochastic manifestation of the ratio of the probabilities of the corresponding events. For objectivity, the responses made by the objects must not be influenced by each other and so must be independent. The probabilities corresponding to the frequencies in Figure 1 are specified in Figure 2. It is the counts in all 4 cells of

Figure 1 which enable the probabilities in Figure 2 to be estimated. Thus the performance of On and Om can be defined in terms of the probabilities corresponding to F10 and F01, giving

$$\frac{\text{Performance of On}}{\text{Performance of Om}} = \frac{Pni1*Pmi0}{Pni0*Pmi1} \qquad (2)$$

	Object On 0	Object On 1
Object Om 0	Pni0*Pmi0	Pni1*Pmi0
1	Pni0*Pmi1	Pni1*Pmi1

Figure 2. The probabilities of the outcome of the administration of agent Ai to objects On and Om. Pni1 is the probability of object On responding to agent Ai successfully.

By specific objectivity, the comparison of On and Om must be independent of which particular agent is used in making the comparison. Therefore the comparison of their performance must be the same for agent Ai and for agent Aj. Therefore

$$\frac{\text{Performance of On}}{\text{Performance of On}} = \frac{Pni1*Pmi0}{Pni0*Pmi1} = \frac{Pnj1*Pmj0}{Pnj0*Pmj1} \qquad (3)$$

Rearranging the terms,

$$\frac{Pni1}{Pni0} = \frac{Pnj1}{Pnj0} * \frac{Pmi1}{Pmi0} * \frac{Pmj0}{Pmj1} \qquad (4)$$

This is true for all m, n, i, j. Thus it is also true if object Om is object O0, whose measure is defined to be at the local origin of the object measurement scale, and also true for agent A0, whose calibration is defined to be at the local origin of the agent measurement scale. Then

$$\frac{Pni1}{Pni0} = \frac{Pn01}{Pn00} * \frac{P0i1}{P0i0} * \frac{P000}{P001} \qquad (5)$$

But, P000/P001 is a constant term dependent on the relative placement of the local origins of the object and agent sub-scales, so that, if they are defined to coincide, then P000 = P001 = 0.5 and P000/P001 = 1.

Then, once the local origin is defined, Pn01/Pn00 is a ratio dependent only on object On, so that log(Pn01/Pn00) can be expressed as Bn. Similarly, log(P0i0/P0i1) is a ratio dependent only on agent Ai, which can be expressed as Di. Equation (5) thus becomes, on taking logarithms and reparameterizing Pni1 to be Pni, so that Pni0 is 1 - Pni,

$$\log(Pni/(1-Pni)) = Bn - Di \tag{6}$$

which is the Rasch model for dichotomous, two-faceted data.

III. Generalizing to many-faceted models using a three-faceted example.

The derivation of a particular form of the many-faceted model demonstrates the general principles by which any other particular form of the many-faceted model can also be derived. The particular model to be derived here is applicable to a three-faceted test in which each judge of a panel of judges awards a rating to each examinee on each item.

Consider the performance of two examinees, On and Om, as rated by a particular judge, Jj, on replications of the same item, Ai. In whatever way the ratings were originally recorded, they have been recoded into K+1 categories ordinally numbered from 0 to K, with each higher numbered category representing a higher level of perceived performance, and with each category having a non zero probability of occurrence.

Categories	Examinee On k	1
Examinee Om k	Fkk	Flk
1	Fkl	Fll

Figure 3. Fkl represents the count of the number of times that examinee On is awarded rating k and examinee Om is rated a 1 by judge Jj across replications of item Ai, where k > 1.

The administration of the numerous replications of item Ai is the "test". The performance levels of examinees On and Om can be compared by their relative frequencies of being rated in the various categories of the rating scale. Following the procedure in the discussion of objectivity, let us summarize part of their performance by a 2x2 cross-tabulation of counts of ratings in categories k and 1 of the rating scale, chosen so that category k is numerically greater than category 1 and so represents a higher performance level. This is depicted in Figure 3.

When both examinees are given the same rating, (Fkk times for a rating of k, and Fll times for a rating of 1), their performance levels are indistinguishable. When the examinees are rated differently (Fkl and Flk

times), the examinee with the greater relative frequency of ratings in category k, the higher category, is perceived to have the higher ability. In comparing performance levels, we intend that the numeric result be independent of the number of replications. Thus, if the test were to be repeated again, and were of the same length, we would expect to get approximately the same numeric result. Moreover, if the two tests were then to be concatenated, we would again expect to obtain about the same result. The division of the two frequencies, Fkl and Flk, is compatible with this expectation because we expect this ratio to be about the same when the test is repeated, and also when the two tests are concatenated. Consequently, the comparative levels of performance of examinees On and Om can be identified by the stated ratio, Fkl/Flk.

$$\frac{\text{Performance level of On}}{\text{Performance level of Om}} \approx \frac{Fkl}{Flk} \qquad (7)$$

In the limit, the ratio of the empirically observed frequencies, Fkl/Flk, becomes the ratio of the probabilities, Pkl/Plk, where Pkl is the probability of examinee On being given a rating of k and examinee Om a rating of l on one replication of item i, and Plk is similarly defined. Thus we define the ratio Pkl/Plk to be the ratio of the examinee's performances.

$$\frac{\text{Performance level of On}}{\text{Performance level of Om}} = \frac{Pkl}{Plk} \qquad (8)$$

But, for objectivity, the ratings given examinees Om and On must be independently awarded by the judge. Consequently,

$$Pkl = Pnijk * Pmijl \qquad (9)$$

and

$$Plk = Pnijl * Pmijk \qquad (10)$$

where Pnijk is the probability of examinee On being given a rating of k on item Ai by judge Jj, and Pnijl, Pmijk, Pmijl are similarly defined. Then

$$\frac{\text{Performance level of On}}{\text{Performance level of Om}} = \frac{Pkl}{Plk} = \frac{Pnijk}{Pnijl} * \frac{Pmijl}{Pmijk} \qquad (11)$$

However, also for objectivity, the relative performance of examinees On and Om must be independent of which particular item is used to compare them. Thus, though performance levels are initially defined in terms of item Ai, the relative performance levels must have the same value when defined in terms of any conceptually equivalent item Ai'. That is

$$\frac{\text{Performance level of On}}{\text{Performance level of Om}} = \frac{Pnijk}{Pnijl} * \frac{Pmijl}{Pmijk} = \frac{Pni'jk}{Pni'jl} * \frac{Pmi'jl}{Pmi'jk} \qquad (12)$$

then

$$\frac{Pnijk}{Pnijl} = \frac{Pmijk}{Pmijl} * \frac{Pni'jk}{Pni'jl} * \frac{Pmi'jl}{Pmi'jk} \qquad (13)$$

For objectivity, this ratio of the probabilities of examinee On being rated in categories k and l must be independent of whichever examinee Om is used in the comparison. So, let us consider examinee O0 with performance level at the local origin of the ability scale. Similarly the ratio must also be independent of whichever item Ai' is used for the comparison. Thus it must also be hold for item A0 chosen to have difficulty at the local origin of the item scale.

$$\frac{Pnijk}{Pnijl} = \frac{P0ijk}{P0ijl} * \frac{Pn0jk}{Pn0jl} * \frac{P00jl}{P00jk} \qquad (14)$$

If, instead of comparing performance levels by means of items Ai and Ai', we compare performance levels by means of the ratings given by judges Cj and Cj' over numerous replications of item Ai, then again we expect the relative performance levels of the examinees to be maintained.

$$\frac{\text{Performance level of On}}{\text{Performance level of Om}} = \frac{Pnijk}{Pnijl} * \frac{Pmijl}{Pmijk} = \frac{Pnij'k}{Pnij'l} * \frac{Pmij'l}{Pmij'k} \qquad (15)$$

so that

$$\frac{Pnijk}{Pnijl} = \frac{Pmijk}{Pmijl} * \frac{Pnij'k}{Pnij'l} * \frac{Pmij'l}{Pmij'k} \qquad (16)$$

Again this must be true if judge Cj' is chosen to be judge C0 with severity at the local origin of the severity scale, and examinee Om is examinee O0, and when item Ai is replaced by item A0. Therefore

$$\frac{Pn0jk}{Pn0jl} = \frac{P00jk}{P00jl} * \frac{Pn00k}{Pn00l} * \frac{P000l}{P000k} \qquad (17)$$

Furthermore, for objectivity, the relative severity levels of judges Cj and Cj' must be maintained whether the judging takes place over numerous replications of the administration of either item Ai or item Ai' to the same examinee On.

$$\frac{\text{Severity level of Cj'}}{\text{Severity level of Cj}} = \frac{Pnijk}{Pnijl} * \frac{Pni'jl}{Pni'jk} = \frac{Pnij'k}{Pnij'l} * \frac{Pni'j'l}{Pni'j'k} \qquad (18)$$

then

$$\frac{Pnijk}{Pnijl} = \frac{Pni'jk}{Pni'jl} * \frac{Pnij'k}{Pnij'l} * \frac{Pni'j'l}{Pni'j'k} \qquad (19)$$

Again this must be true if judge Cj' is judge C0 chosen at the origin of the severity scale, and examinee On is examinee O0, and item Ai' is item A0.

$$\frac{P0ijk}{P0ijl} = \frac{P00jk}{P00jl} * \frac{P0i0k}{P0i0l} * \frac{P0001}{P000k} \qquad (20)$$

Substituting (20) and (17) in (14), and simplifying

$$\frac{Pnijk}{Pnijl} = \frac{Pn00k}{Pn001} * \frac{P0i0k}{P0i01} * \frac{P00jk}{P00j1} * \left(\frac{(P00001)^2}{(P0000k)}\right) \qquad (21)$$

which gives a general form in which each term is an expression of the relationship between a component of a facet and the local origin of a subscale, in the context of a particular pair of categories.

IV. The three-faceted dichotomous model.

Considering (21) as a dichotomous model where k=1 ("right") and l=0 ("wrong"), then this equation expresses the ratio of the probabilities of the possible outcomes as a product of terms which relate each component of each facet with the local origin of its subscale. These terms are independent of whichever other examinees, items and judges are included in the test situation. To consider these terms in an additive way, we can take logarithms and assign a numerical direction to each term in accordance with conventional interpretation. Let

Bn = log(Pn001/Pn000), which is defined to be the ability of an examinee
Di = log(P0i00/P0i01), which is defined to be the difficulty of an item
Cj = log(P00j0/P00j1), which is defined to be the severity of a judge

We also define the relationship of the subscales, such that the probability of "original" examinee O0 being rated a "1" by judge C0 on item A0 is 0.5, so that the last term of equation (21) becomes 0.

Reparameterizing Pnij = Pnij1, so that 1 - Pnij = Pnij0, gives equation (22), the three-faceted Rasch model for the dichotomous case.

$$\log(Pnij/(1-Pnij)) = Bn - Di - Cj \qquad (22)$$

V. The three-faceted rating scale model.

When we consider examinees On and Om in the more general circumstances of a rating scale, we do not wish the comparison of their abilities to depend on which particular pair of categories of the rating scale are used for the

comparison. So we return to equation (11) which stated:

$$\frac{\text{Performance level of On}}{\text{Performance level of Om}} = \frac{Pkl}{Plk} = \frac{Pnijk}{Pnijl} * \frac{Pmijl}{Pmijk} \qquad (11)$$

We wish to generalize this equation to any pair of categories, but the rating scale categories are not independent but structured. In order to determine the structure in an objective manner, we require that performance levels are invariant when they are compared using any pair of adjacent categories in ascending order. This is the only possible objective structuring since invariance, which is not over adjacent categories, but over some pairing of non-adjacent categories results in a contradiction or indeterminacy in the rating scale structure. Thus, if a rating scale has 3 categories and performance levels are to be invariant only when the top and bottom categories are used for the comparison in (11), then performance levels based on the middle category are indeterminate, and are not objective.

Invariance in relative performance when categories are chosen such that k is one greater than l, and also k' is chosen one greater than l', yields

$$\frac{\text{Performance level of On}}{\text{Performance level of Om}} = \frac{Pnijk}{Pnijl} * \frac{Pmijl}{Pmijk} = \frac{Pnijk'}{Pnijl'} * \frac{Pmijl'}{Pmijk'} \qquad (23)$$

so that

$$\frac{Pnijk}{Pnijl} = \frac{Pnijk'}{Pnijl'} * \frac{Pmijk}{Pmijl} * \frac{Pmijl'}{Pmijk'} \qquad (24)$$

Since we want this result to be generalizable, we must be able to substitute examinee O0, item A0, and judge C0,

$$\frac{Pn00k}{Pn00l} = \frac{Pn00k'}{Pn00l'} * \frac{P000k}{P000l} * \frac{P000l'}{P000k'} \qquad (25)$$

Reordering the terms,

$$\frac{Pn00k}{Pn00l} = \left(\frac{Pn00k'}{Pn00l'} * \frac{P000l'}{P000k'}\right) * \frac{P000k}{P000l} \qquad (26)$$

the two terms in parentheses are invariant over changes in choice of pairs of categories and so are independent of the local structure of the rating scale, but they are not independent of the choice of object, so we can accordingly rewrite them as Pn00, so that

$$\frac{Pn00k}{Pn00l} = Pn00 * \frac{P000k}{P000l} \qquad (27)$$

Similar equations hold for P0i0k/P0i01 and P00jk/P00j1, so that, substituting into (21),

$$\frac{Pnijk}{Pnij1} = Pn00 * P0i0 * P00j * \frac{P000k}{P0001} \qquad (28)$$

We have an equation in which the ratio of the probabilities of particular outcomes is the product of terms which depend only on a single component and the local origin of its subscale, combined with a term dependent on the pair of categories used for the comparison. Let

Bn = log(Pn00), which is defined to be the examinee ability,
Di = - log(P0i0), which is defined to be the item difficulty,
Cj = - log(P00j), which is defined to be the judge severity
Fk = - log(P000k/P0001), which is defined to be the difficulty of the step from category k-1 (=1) to category k of the rating scale.

Then equation (28), the three-faceted Rasch rating scale model becomes

$$\log(Pnijk/Pnijk-1) = Bn - Di - Cj - Fk \qquad (29)$$

In (29), the parameters relating to the particular examinees, items, and judges interacting to make each rating have been separated, and so (29) is objective in that the parameters of the particular examinees, judges and items enter independently into the rating process. Nevertheess, the parameter estimate of Bn, say, could be inflated by an arbitrary amount so long as the other parameter estimates were deflated accordingly. Thus the actual placement of the examinee, item, judge and step subscales within the common frame of reference is arbitrary. By convention, local origins for each subscale are chosen such that the mean calibrations of the items, of the judges, and of the rating scale steps are each zero. The local origin of the examinee abilities' subscale is then defined uniquely by the model.

VI. Conclusion

The manner in which the dichotomous two-faceted Rasch model can be derived from objectivity has been demonstrated to extend to an example of the many-faceted model which includes a rating scale. Any other particular form of the many-faceted model may also be derived in a similar way.

Bibliography:
Rasch, G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, 1960, and Chicago: University of Chicago Press. 1980.

Rasch. G. Objective Comparisons. UNESCO Seminar, Voksenasen, Oslo. 1964

Rasch, G. An individual-centered approach to item analysis with two categories of answers. Unpublished paper. 1966.

Wright, B. D. & Linacre, J. M. Rasch Models derived from objectivity. Rasch SIG Newsletter, 1987, 1:1.